

## AND-Type versus WITH-Type Conjunctions: Towards a Corpus-Based Study

*Beata Trawiński (University of Vienna)*

### 1. Introduction

In this paper, I discuss two strategies for expressing nominal conjunction: the comitative preposition *with* and the ordinary conjunction *and*. These two strategies are available in many languages of the world, and they have been widely discussed in linguistic literature, being examined from different theoretical perspectives. The sentences in (1) and (2) provide examples of an AND-type and a WITH-type conjunction in Polish, respectively.

(1) Adam     i     Monika     wyszli.  
Adam<sub>Nom</sub> and Monika<sub>Nom</sub> left<sub>Pl</sub>  
'Adam and Monika left.'

(2) Adam     z     Moniką     wyszli.  
Adam<sub>Nom</sub> with Monika<sub>Instr</sub> left<sub>Pl</sub>  
'Adam and Monika left.'

Both sentences contain the nouns *Adam* and *Monika*. In (1), these nouns both occur in the nominative case and are connected by the conjunction *i* 'and'. In (2), they are connected with the comitative preposition *z* 'with'. Here, while the noun *Adam* bears the nominative case, *Monika* is assigned the instrumental case. (1) and (2) clearly differ in their form but they seemingly mean the same, which is indicated by the English translations. Both sentences describe an event of leaving in which both Adam as well as Monika participate.

The question addressed in this paper is whether these two types of conjunctions are indeed fully equivalent or whether there are any factors determining the choice between the two forms. The answer to this question is interesting and relevant from the empirical, theoretical and practical point of view, the last one being that of the learners of Polish. I will try to answer this question by looking at Polish data and providing corpus evidence.

In the next section, I make a brief overview of previous discussions on (non)equivalence of these two ways for expressing conjunction. In Section 3, I present an analysis of the distribution of AND-type and WITH-type conjunctions containing relational and non-relational nouns in the National Corpus of

Polish. Based on the results of this analysis, I propose in Section 4 a new approach to relatedness for Polish. In Section 5, I discuss the distribution of AND- and WITH-type conjunctions in different registers and channels. Finally, in Section 6, I sum up the discussion and make suggestions for future research.

## 2. Previous discussions

There are two main approaches to the relationship between AND-type and WITH-type conjunctions found in linguistic literature. According to the first one, held by Schwartz (1985, 1988b,a) and Dalrymple et al. (1998), these two types of constructions are semantically fully equivalent. Schwartz investigates AND-type and WITH-type coordinations involving (dropped) plural pronouns in many languages, including Spanish, Hungarian, Russian and Polish, and suggests a conjunctive interpretation of both types. She also assumes that in both types of constructions, the involved nouns are assigned the same Theta-role. Similarly, Dalrymple et al. (1998) discuss AND-type and WITH-type conjunctions in Russian, and also argue for a denotational uniformity of these constructions, which are analyzed there as first-order sums of individuals.

By contrast, Miller (1971), Camacho (1994, 2000), Kopcińska (1995), McNally (1993) and Urtz (1994) claim that there are some differences between AND-type and WITH-type conjunctions. Miller (1971) claims that in contrast to individuals in the denotation of ordinary conjunction, individuals in the denotation of WITH-type conjunction are related to each other in some sense. Given this, he analyzes Russian WITH-type conjunction as a close type of coordination and nominal AND-type conjunction as a loose type of coordination. Interrelatedness of individuals referred to by WITH-type conjunctions has also been postulated by Camacho (1994, 2000) for Spanish and Kopcińska (1995) for Polish, who analyzes this relationship in terms of possession. A similar observation has also been made in McNally (1993), who discusses Russian and Polish data and suggests an account of the relatedness of individuals in WITH-type conjunctions based on conventional implicatures. Finally, indications of a close relatedness of individuals in the denotation of WITH-type conjunctions can be found in Urtz (1994), who investigates Russian data. However, she claims that this relatedness does not necessarily imply a spatiotemporal association.

Given the examples in (1) and (2), the theories of relatedness predict that Adam and Monika in (2) are related to each other, for example as husband and wife, brother and sister, or just close friends. As for (1), these theories would predict that there is no such close relationship between Adam and Monika. In the following section, I demonstrate that these predictions are supported by corpus-data, but only to some extent.

### 3. Corpus evidence

For the present study, I used the National Corpus of Polish (in Polish: Narodowy Korpus Języka Polskiego, NKJP), freely available at <http://www.nkjp.pl> and described in Przepiórkowski et al. (2008, 2009, 2010), among others. The NKJP is annotated morphosyntactically, includes bibliographic metadata and information about channels, i.e. the publication medium (such as press, book, Internet, spoken, announcements and manuscripts), and contains various types of texts, such as fiction and non-fiction literature, journalistic texts, academic and instructive writing, text- and guidebooks, texts acquired from the Internet (including forums, chat rooms, mailing lists and static websites), spoken and quasi-spoken texts. I used a version of May 2010 with about 1075 million segments.

#### 3.1 The procedure

I selected a number of nouns which are inherently relational, just due to their lexical meaning. A typical instance of such relational nouns are kinship terms such as *mother*, *father*, *child*, *son*, *daughter* etc. Also, certain profession terms are often used relationally, for example *doctor*, *lawyer*, *chauffeur* or *secretary*. If a noun of this kind appears with another noun within an AND-type or a WITH-type conjunction, then it can be assumed that the referents of these two nouns are somehow related to each other. Thus, by using relational nouns in combination with the conjunction *i* or with the preposition *z* and with another arbitrary noun, AND-type and WITH-type conjunctions fulfilling the criterion of relatedness can be extracted.

For searching the NKJP, I used the search engine Poliquarp (Janus & Przepiórkowski 2007), which involves a very expressive query language. I entered the following query to identify WITH-type conjunctions including relational nouns, where RN stands for the orthographic form of a relational noun:

(3) [pos=subst & case=nom & !base="spotkanie|rozmowa"] "z|ze" [] {0,1}  
[orth=RN & case=inst]

Thus, I looked for sequences consisting of a noun in the nominative case, the preposition *z* or its vocalic alternation *ze*, possibly another expression, and a relational noun in the instrumental case. Note that I excluded two nouns from my search: *spotkanie* 'meeting' and *rozmowa* 'talk'. These nouns are themselves relational nouns selecting PPs headed by the preposition *z* as their arguments. This preposition has the same form as the comitative preposition *z*, but it bears a different meaning. Since these two nouns are used quite frequently in Polish, not excluding them from the search would have skewed the results. The results of the query in (3) include expressions such as *matka z dzieckiem* 'a mother and her

child' or *Clinton ze swoim kierowcą* 'Clinton and his chauffeur'.

To identify AND-type conjunctions, I used the query in (4). In this case, I looked for sequences consisting of a noun in nominative, the conjunction *i* or *oraz*, possibly another expression, and a relational noun in nominative.

- (4) [pos=subst & case=nom] "i|oraz" [] {0,1} [orth=RN & case=nom]

The examples of the results are *mąż i żona* 'husband and wife' and *Sławomir i jego ojciec* 'Sławomir and his father'.

I entered the above queries for about 30 kinship terms and for about 50 terms referring to a profession. Unfortunately, many of the expressions found were not frequent enough for evaluation, but some of the results could be adequately analyzed. Those expressions include the following relational nouns:

- (5) *dziecko* 'child', *mąż* 'husband', *żona* 'wife', *córka* 'daughter', *ojciec* 'father', *brat* 'brother', *syn* 'son', *matka* 'mother', *przewodnik* 'guide', *sekretarka* 'secretary', *przełożony* 'principal', *kierowca* 'chauffeur', *lekarz* 'doctor', *wspólnik* 'collaborator', *pielęgniarka* 'nurse', *adwokat* 'lawyer', *szef* 'boss'

My key assumption was that if the hypothesis of relatedness is true, then the frequency of the relational nouns in (5) should be significantly higher in WITH-type conjunctions than in AND-type conjunctions (relative to the total frequency of AND-type and WITH-type conjunctions involving these nouns).

### 3.2 The association measure

To examine this hypothesis, I used a very straightforward association measure that I will call *occurrence ratio* (OR). It is computed for a relational noun R using its overall frequency N in the two conjunction types as defined in (6) and (7).

$$(6) \text{OR}_{\text{WITH}} := \frac{N_{\text{WITH}}}{N_{\text{WITH}} + N_{\text{AND}}}$$

$$(7) \text{OR}_{\text{AND}} := \frac{N_{\text{AND}}}{N_{\text{WITH}} + N_{\text{AND}}}$$

$\text{OR}_{\text{WITH}}$  reflects the fraction of the occurrence of WITH-type conjunctions involving a relational noun R to the sum of this occurrence and the occurrence of AND-type conjunctions involving R.  $\text{OR}_{\text{AND}}$  reflects the fraction of the occur-

rence of AND-type conjunctions involving a relational noun R to the sum of this occurrence and the occurrence of WITH-type conjunctions involving R. Both OR values range between 0 and 1. The theory of relatedness predicts that  $OR_{WITH}$  is greater than  $OR_{AND}$ .

### 3.3 The results

I first calculated the OR values for the 8 kinship terms; see Table 1.

Table 1: The occurrence ratios for kinship terms

Lemma	WITH		AND		Total	
	$N_{WITH}$	$OR_{WITH}$	$N_{AND}$	$OR_{AND}$	$N_{WITH} + N_{AND}$	OR
dziecko 'child'	2070	0.73	763	0.27	2822	1.0
mąż 'husband'	459	0.51	436	0.49	859	1.0
żona 'wife'	1052	0.50	1065	0.50	2117	1.0
córka 'daughter'	455	0.39	709	0.61	1164	1.0
ojciec 'father'	494	0.35	907	0.65	1401	1.0
brat 'brother'	310	0.33	622	0.67	932	1.0
syn 'son'	540	0.29	1317	0.71	1857	1.0
matka 'mother'	431	0.25	1315	0.75	1746	1.0

The total frequency of the noun *dziecko* 'child' in WITH-type conjunctions is 2070, its total frequency in AND-type conjunctions is 763, and its total frequency in both constructions is 2822. On the basis of these frequencies, I calculated the  $OR_{WITH}$  value for this noun, which is 0.73, and its  $OR_{AND}$  value, which is 0.27. In this case, the  $OR_{WITH}$  value is significantly higher than the corresponding  $OR_{AND}$  value ( $\chi^2 = 317.885$  (df = 1),  $p < 0.0001$ ), and this is exactly what the theory of relatedness predicts.

However, if the OR values of the other examined kinship terms are considered, this result appears to be a single case. As can be seen in Table 1, the  $OR_{WITH}$  and  $OR_{AND}$  values for the nouns *mąż* 'husband' and *żona* 'wife' are equal, and for the remaining nouns, the  $OR_{AND}$  values are even higher (in part significantly) than the corresponding  $OR_{WITH}$  values.

Exactly the same results can be observed for relational nouns referring to a profession. These results are presented in Table 2. Here, all  $OR_{AND}$  values are higher than the  $OR_{WITH}$  values, ranging between 0.61 and 0.82.

Table 2: The occurrence ratios for profession terms

Lemma	WITH		AND		Total	
	$N_{WITH}$	$OR_{WITH}$	$N_{AND}$	$OR_{AND}$	$N_{WITH} + N_{AND}$	OR
przewodnik 'guide'	91	0.39	141	0.61	232	1.0
sekretarka 'secretary'	22	0.37	37	0.63	59	1.0
przełożony 'principal'	15	0.36	27	0.64	42	1.0
kierowca 'chauffeur'	115	0.35	213	0.65	328	1.0
lekarz 'doctor'	239	0.34	468	0.66	707	1.0
wspólnik 'collaborator'	13	0.24	42	0.76	55	1.0
pielęgniarka 'nurse'	28	0.23	93	0.77	121	1.0
adwokat 'lawyer'	18	0.20	74	0.80	92	1.0
szef 'boss'	226	0.18	1011	0.82	1237	1.0

To conclude, the distribution of relational nouns in AND-type and in WITH-type conjunctions in the NKJP clearly challenge the theory of relatedness.

### 3.4 Further observations

To additionally check the results presented above, I analyzed the distribution of non-relational nouns in AND-type and WITH-type conjunctions in the NKJP. For this purpose, I used the following nouns:

- (8) *reżyser* 'director', *poeta* 'poet', *pisarz* 'author', *aktor* 'actor', *mechanik* 'mechanic', *informatyk* 'computer scientist', *malarz* 'painter', *wykładowca* 'lecturer'

The results of this examination are presented in Table 3, and they appear to be very interesting.

Table 3: The occurrence ratios for non-relational nouns

Lemma	WITH		AND		Total	
	$N_{WITH}$	$OR_{WITH}$	$N_{AND}$	$OR_{AND}$	$N_{WITH} + N_{AND}$	OR
reżyser 'director'	100	0.08	1160	0.92	1260	1.0
poeta 'poet'	65	0.07	829	0.93	894	1.0
pisarz 'author'	76	0.06	1112	0.94	1188	1.0
aktor 'actor'	54	0.06	793	0.94	847	1.0
mechanik 'mechanic'	5	0.06	75	0.94	80	1.0
informatyk 'computer scientist'	4	0.05	70	0.95	74	1.0
malarz 'painter'	14	0.04	314	0.96	328	1.0
wykładowca 'lecturer'	5	0.02	198	0.98	203	1.0

Table 3 demonstrates that the proportion between the  $OR_{WITH}$  values and the  $OR_{AND}$  values for non-relational nouns considerably differs from that observed

for relational nouns. In particular, the  $OR_{WITH}$  values for non-relational nouns are extremely low: they are always under 0.1, while the corresponding values for relational nouns range between 0.51 and 0.18.

These observations suggest that relatedness indeed plays a role in the usage of WITH-type and AND-type conjunctions. But they also show that relatedness is not a decisive factor for the choice of one of these constructions.

#### 4. Revising the hypothesis of relatedness

On the basis of the corpus data presented above, I propose the following, weak approach of relatedness for Polish:

1. If two nouns are connected with each other, then the use of the conjunction *i* is preferred over the use of the comitative preposition *z*;
2. If the nouns refer to unrelated individuals, this preference is extremely strong;
3. If the nouns refer to related individuals, this preference is considerably weaker.

As can be seen, this approach is formulated in terms of preference rather than requirement. An interesting question that might be addressed here is whether there are any factors, for example the text type or channel, that influence this preference. In the next section, I take a closer look at this issue.

#### 5. Further issues and open questions

To examine whether there are any differences in the distribution of AND-type and WITH-type conjunctions involving relational and non-relational nouns over different registers and channels, I conducted the following experiment.

I first compared the frequencies of AND-type conjunctions containing the relational noun *ojciec* 'father' in different types of texts with the corresponding frequencies of WITH-type conjunctions. Then I did the same for the non-relational noun *pisarz* 'author'. The results are given in Table 4 and Table 5, where  $N$  refers to the frequency of an AND-type or WITH-type conjunction in a given text type, and  $TTR$  is the ratio of this frequency to the overall frequency of this expression type in the corpus.

Table 4: The distribution of AND-type and WITH-type conjunctions involving a relational noun in different text types

Type of Text	WITH		AND	
	N	TTR	N	TTR
fiction	56	0.11	78	0.09
non-fiction	17	0.03	23	0.03
journalism	122	<b>0.25</b>	328	<b>0.36</b>
academic writing	7	0.01	13	0.01
instructive writing	9	0.02	11	0.01
unclassified non-fiction	1	0.00	1	0.00
miscellaneous written texts	0	0.00	1	0.00
Internet	279	<b>0.56</b>	441	<b>0.49</b>
conversational texts	0	0.00	1	0.00
spoken texts from the media	0	0.00	0	0.00
quasi-spoken texts	3	0.01	10	0.01
<b>Total</b>		494		907

Table 5: The distribution of AND-type and WITH-type conjunctions involving a non-relational noun in different text types

Type of Text	WITH		AND	
	N	TTR	N	TTR
fiction	4	0.05	10	0.01
non-fiction	1	0.01	16	0.01
journalism	49	<b>0.64</b>	231	<b>0.21</b>
academic writing	3	0.04	21	0.02
instructive writing	0	0.00	8	0.01
unclassified non-fiction	0	0.00	3	0.00
miscellaneous written texts	0	0.00	2	0.00
Internet	18	<b>0.24</b>	818	<b>0.74</b>
conversational texts	0	0.00	0	0.00
spoken texts from the media	0	0.00	0	0.00
quasi-spoken texts	1	0.01	3	0.00
<b>Total</b>		76		1112

Table 4 demonstrates that the WITH-type and the AND-type conjunctions involving the relational noun *ojciec* 'father' are distributed very similarly over different types of texts. However, Table 5 presents a different situation. According to the results in this table, almost all WITH-type and AND-type conjunctions involving the non-relational noun *pisarz* 'author' occur in journalistic texts and



on the Internet, but the proportions between these occurrences are not parallel, as in Table 4, but reverse.

A similar observation can be made with respect to the distribution of AND-type and WITH-type coordinations in different channels. This is shown in Table 6 and Table 7, where N refers to the frequency of an expression in a given channel, and CR is a ratio of this frequency to the overall frequency of this expression in the corpus.

Table 6: The distribution of AND-type and WITH-type conjunctions involving a relational noun in different channels

Type of Text	WITH		AND	
	N	CR	N	CR
press	121	<b>0.24</b>	340	<b>0.37</b>
book	93	<b>0.19</b>	126	<b>0.14</b>
Internet	280	<b>0.57</b>	441	<b>0.49</b>
spoken	0	0.00	0	0.00
announcements	0	0.00	0	0.00
manuscripts	0	0.00	0	0.00
<b>Total</b>	494		907	

Table 7: The distribution of AND-type and WITH-type conjunctions involving a non-relational noun in different channels

Type of Text	WITH		AND	
	N	CR	N	CR
press	49	<b>0.64</b>	233	<b>0.21</b>
book	8	0.11	61	0.05
Internet	19	<b>0.25</b>	818	<b>0.74</b>
spoken	0	0.00	0	0.00
announcements	0	0.00	0	0.00
manuscripts	0	0.00	0	0.00
<b>Total</b>	76		1112	

Table 6 demonstrates that the WITH-type and AND-type conjunctions including the relational noun *ojciec* 'father' have very similar distribution over different channels. By contrast, the majority of WITH-type conjunctions containing the

non-relational noun *pisarz* 'writer' occur in press and books, whereas over 70 percent of the corresponding AND-type conjunctions appear in Internet texts.

These results suggest that the text type and channel have no impact on the choice between the two conjunction types in the case of relational nouns. As far as non-relational nouns are concerned, the usage of the WITH-type conjunction is preferred in journalistic texts (in terms of the register) and in press and books (in terms of the channel). The AND-type conjunction is used more frequently in Internet texts. It remains to be explained what (pragmatic, stylistic etc.) factors drive these distributional patterns and whether the discrepancies in the distribution over different registers and channels correlate with the disproportion between  $OR_{WITH}$  and  $OR_{AND}$  values for relational and non-relational nouns.

## 6. Summary and outlook

In this paper, I have discussed two types of nominal conjunctions in Polish: the AND-type and the WITH-type conjunction. My objective was to find out whether or not these two kinds of expressions are subject to the same usage conditions. In particular, I tried to verify the hypothesis that a WITH-type conjunction is chosen whenever its referents are related to each other. For this purpose, I have analyzed the distribution of WITH-type and AND-type conjunctions involving selected relational nouns in the National Corpus of Polish. The corpus evidence suggests that the hypothesis of relatedness is correct only to some extent, and so it should be reformulated in terms of preference rather than requirement. The corpus data further show that there is no difference in the distribution of AND-type and WITH-type conjunctions involving relational nouns in different types of texts and different channels. However, AND-type and WITH-type conjunctions involving non-relational nouns have different distribution in various registers and channels. It remains to be explained whether there is a correlation between different distributional patterns of AND-type and WITH-type conjunctions involving relational and non-relational nouns in different registers and channels and the difference in their respective occurrence ratios.

## References

- Camacho, José. 1994. Comitative Coordination in Spanish. In Cl. Parodi, C.s Quicoli, M. Salterelli & Mariá Luisa Zubizarreta (eds.), *Aspects of Romance Linguistics, Selected Papers from the Linguistic Symposium on Romance Languages*. Washington: Georgetown University Press. 107–122.
- Camacho, José. 2000. Structural Restrictions on Comitative Coordination. *Linguistic Inquiry* 31, 366–375.
- Dalrymple, Mary, Irene Hayrapetian & Tracy Holloway King. 1998. The Semantics of the Russian Comitative Construction. *Natural Language and Linguistic Theory* 16,

597–631.

- Janus, Daniel & Adam Przepiórkowski. 2007. Poliqarp: An open source corpus indexer and search engine with syntactic extensions. In *Proceedings of ACL 2007 Demo Session*.
- Kopcińska, Dorota. 1995. Czy słowo z może pełnić taką samą funkcję jak słowo i?. In M. Grochowski (ed.), *Wyrażenia funkcyjne w systemie i tekście*. Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika. 125–136.
- McNally, Louise. 1993. Comitative Coordination: A Case Study in Group Formation. *Natural Language and Linguistic Theory* 11, 347–379.
- Miller, Jim. 1971. Some Types of 'Phrasal Conjunction' in Russian. *Journal of Linguistics* 8, 55–69.
- Przepiórkowski, Adam, Rafał L. Górski, Marek Łaziński & Piotr Pęzik. 2009. Recent Developments in the National Corpus of Polish. In J. Levická & R. Garabík (eds.), *NLP, Corpus Linguistics, Corpus Based Grammar Research: Proceedings of the Fifth International Conference*. Brno: Tribun. 302–309.
- Przepiórkowski, Adam, Rafał L. Górski, Marek Łaziński & Piotr Pęzik. 2010. Recent Developments in the National Corpus of Polish. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation, LREC2010*, Valletta, Malta. ELRA.
- Przepiórkowski, Adam, Rafał L. Górski, Barbara Lewandowska-Tomaszczyk & Marek Łaziński. 2008. Towards the National Corpus of Polish. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation, LREC2008*. Marrakech: ELRA. 827–830.
- Schwartz, Linda. 1985. Plural Pronouns, Coordination and Inclusion. In N. Stenson (ed.), *Papers from the Tenth Minnesota Regional Conference on Language and Linguistics*. Minneapolis: University of Minnesota. 152–184.
- Schwartz, Linda. 1988a. Asymmetric Feature Distribution in Pronominal Coordinations. In M. Barlow & C. A. Ferguson (eds.), *Agreement in Natural Language: Approaches, Theories, Descriptions*. Stanford: CSLI. 237–249.
- Schwartz, Linda. 1988b. Conditions for Verb-Coded Coordinations. In M. Hammond, E. Moravcsik & J. Wirth (eds.), *Studies in Syntactic Typology, Typological Studies in Language (TSL)*. Amsterdam: John Benjamins. 53–73.
- Urtz, Bernadette. 1994. The Syntax, Semantics and Pragmatics of a Nominal Conjunction. The Case of Russian "S". Doctoral dissertation, Harvard University.